

The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements

Abstract Complete landslide inventories are rarely available. The objectives of this study were to (i) elaborate the influence of incomplete landslide inventories on statistical landslide susceptibility models and to (ii) propose suitable modelling strategies that can reduce the effects of inventory-based incompleteness. In this context, we examined whether the application of a novel statistical approach, namely mixed-effects models, enables predictions that are less influenced by such inventory-based errors.

The study was conducted for (i) an area located in eastern Austria and (ii) a synthetically generated data set. The applied methodology consisted of a simulation of two different inventory-based biases and an in-depth evaluation of subsequent modelling results. Inventory-based errors were simulated by gradually removing landslide data within forests and selected municipalities. The resulting differently biased inventories were introduced into logistic regression models while we considered the effects of including or excluding predictors that are directly related to the respective inventory-based bias. Mixed-effects logistic regression was used to account for variation that was due to an inventory-based incompleteness.

The results show that most erroneous predictions, but highest predictive performances, were obtained from models generated with highly incomplete inventories and predictors that were able to directly describe the respective incompleteness. An exclusion of such bias-describing predictors led to systematically confounded relationships. The application of mixed-effects models proved valuable to produce predictions that were least affected by inventory-based errors.

This paper highlights that the degree of inventory-based incompleteness is only one of several aspects that determine how an inventory-based bias may propagate into the final results. We propose a four-step procedure to deal with incomplete inventories in the context of statistical landslide susceptibility modelling.

Keywords Landslide susceptibility · Statistical modelling · Landslide inventory · Mixed-effects model · Error propagation · Validation · Logistic regression

Introduction

Landslide inventories are commonly compiled to investigate the geomorphic evolution of steeplands or to analyze earth surface dynamics (Soeters and Van Westen 1996; Hovius et al. 1997; Muenchow et al. 2012). Spatial information on past slope movements is also used to calibrate and validate empirically based landslide susceptibility models (Brenning 2005; Van Westen et al. 2008; Guzzetti et al. 2012; Petschko et al. 2016). The derived landslide susceptibility maps regularly serve as a basis for decision-making in hazard prevention and spatial planning (Fell et al. 2008; Greiving et al. 2012; Guillard and Zezere 2012; Corominas et al. 2013; Petschko et al. 2014b) because they depict

locations where landslides are more or less likely to occur in the future (Brabb 1984; Guzzetti et al. 2006b). Landslide susceptibility maps do not provide information on ‘when’ a landslide will occur or ‘how large’ or ‘intensive’ a future landslide will be (Guzzetti et al. 2006a).

In principle, statistical landslide susceptibility models are generated by relating spatial information on past landslide activities (i.e. landslide presence/absence) to static geoenvironmental factors (e.g. topography, lithology) using statistical or machine-learning techniques. The generated empirical relation, commonly expressed as a relative susceptibility score, is then applied to each spatial unit of an area (e.g. grid cell, slope unit) (Rossi et al. 2010; Reichenbach et al. 2014; Goetz et al. 2015). The validity and generalizability of such spatial predictions are commonly assessed by interpreting inventory-based predictive performance estimates (Brenning 2005; Guzzetti et al. 2006b; Frattini et al. 2010). Since metrics such as the Area under the Receiver Operating Characteristic Curve (AUROC) or the area under the prediction rate curve are calculated for one or multiple independent landslide test samples (i.e. samples not used to train the model), they are frequently considered to summarize the capability of a predictive model to identify landslide-prone areas (Chung and Fabbri 2003; Remondo et al. 2003; Beguería 2006; Guzzetti et al. 2006b). Such inventory-based metrics are also taken into account to evaluate different classification algorithms (Goetz et al. 2015; Steger et al. 2016a) or the utility of specific predictor combinations (Iovine et al. 2014; Conoscenti et al. 2016), the spatial transferability of modelling results (Petschko et al. 2014b; Lombardo et al. 2014), the influence of sample sizes (Petschko et al. 2014b; Heckmann et al. 2014; Hussin et al. 2016), the effect of sampling strategies (Regmi et al. 2014; Conoscenti et al. 2016; Hussin et al. 2016) or the impact of data set qualities (Galli et al. 2008; Fressard et al. 2014).

A number of studies outline that a reliable landslide inventory is a vital component to achieve high-quality statistical landslide susceptibility models, also because most analysis steps are dependent on a correct representation of past landslide occurrences (Guzzetti et al. 2006b; Cascini 2008; Fell et al. 2008; Harp et al. 2011; Petschko et al. 2014b; Steger et al. 2016a, b). Of particular concern are the positional accuracy and completeness of landslide information (Malamud et al. 2004; Chacón et al. 2006; Galli et al. 2008; Guzzetti et al. 2012; Xu et al. 2015; Petschko et al. 2016; Santangelo et al. 2015). The positional accuracy of an inventory is reliant on, e.g. the type and quality of the available mapping basis, time availability and the specific characteristics of landslides and the study site (Ardizzone et al. 2002; Harp et al. 2011; Guzzetti et al. 2012; Petschko et al. 2016; Santangelo et al. 2015). Even though modern technologies such as differential Global Positioning Systems (GPS) and Light Detection and Ranging (LiDAR) may facilitate a positionally precise localization of visible landslide features, complete landslide inventories may still be challenging to achieve.

In agricultural land or near transportation infrastructure, for instance, terrain features indicative of landslide activity may more frequently and more quickly be blurred or removed by human activities (e.g. remediation work, planation). This may favour an overrepresentation of landslides within forest areas and an underrepresentation in agricultural land (Bell et al. 2012; Petschko et al. 2016; Conoscenti et al. 2016). In contrast, inventories mapped by visually analyzing multi-temporal aerial photographs may be substantially incomplete within forests because treetops may 'hide' a considerable portion of geomorphic features (Brardinoni et al. 2003; Jacobs et al. 2016). Landslide inventories compiled from public reports may overrepresent landslides in closer proximity to infrastructure and underrepresent slope movements in remote and forest areas (Guzzetti et al. 1999; Steger et al. 2016a).

Several studies compared statistical landslide susceptibility models produced from heterogeneous inventories (Ardizzone et al. 2002; Galli et al. 2008; Zêzere et al. 2009; Fressard et al. 2014; Steger et al. 2015, 2016a). However, a differentiated evaluation of the propagation of potential inventory-based errors into landslide susceptibility models was hampered due to the practical inseparability of positional accuracy and inventory completeness as well as the lack of truly accurate reference inventories. While a previous study examined the effects of positional accuracy in detail (cf. Steger et al. 2016b), the present follow-up study devotes particular attention to the second quality-defining criterion, namely the completeness of a landslide inventory.

This research focuses on the impact of a spatially heterogeneous completeness of landslide information on statistical landslide susceptibility models by artificially introducing two different mapping biases into (i) an available landslide inventory and (ii)

synthetically generated data sets. The main objective was to examine the influence of systematically incomplete landslide inventories on modelled relationships and validation results. An additional goal was to propose suitable modelling strategies that can mitigate the effects of incompleteness. In this context, we built upon earlier findings (Steger et al. 2016a) and explored for the first time the potential of mixed-effects models to tackle the problem of confounded relationships as a result of inventory-based biases.

Study area

The study area (20 km × 5 km) belongs to the districts of Amstetten and Waidhofen/Ybbs in the western part of the federal state of Lower Austria (Fig. 1d). The prevalent undulating landscape of the Flysch Zone (81 km²; mean slope, 12.3°), with its intensively weathered alternating sediment sequences, is highly susceptible to landslides of the slide-type movement. The less steep northern portion of the study area is partly covered by clastic sediments of the Molasse Zone (3 km²; mean slope, 5.1°) while the valley floors are mainly covered by quaternary sediments (16 km²; mean slope, 5°) (Fig. 1a).

The hilly parts located in higher elevations are intensively used for cattle farming (pastures, 50 km²) while arable land (20 km²) is dominant in the lowlands. Forest areas (26 km²) are predominantly located in the hilly parts of the Flysch Zone (Fig. 1c). Settlements and main roads are primarily located within the valley floors while single farms, and smaller roads are also scattered over the steeper hilly areas. In total, built-up areas account for 4 km² (Eder et al. 2011).

The study site experiences oceanic climate influences from the West and continental influences from the East. Mean annual precipitation amounts generally increase with elevation and range from 900 mm in the North (elevation, 350 m a.s.l.) to 1100 mm in

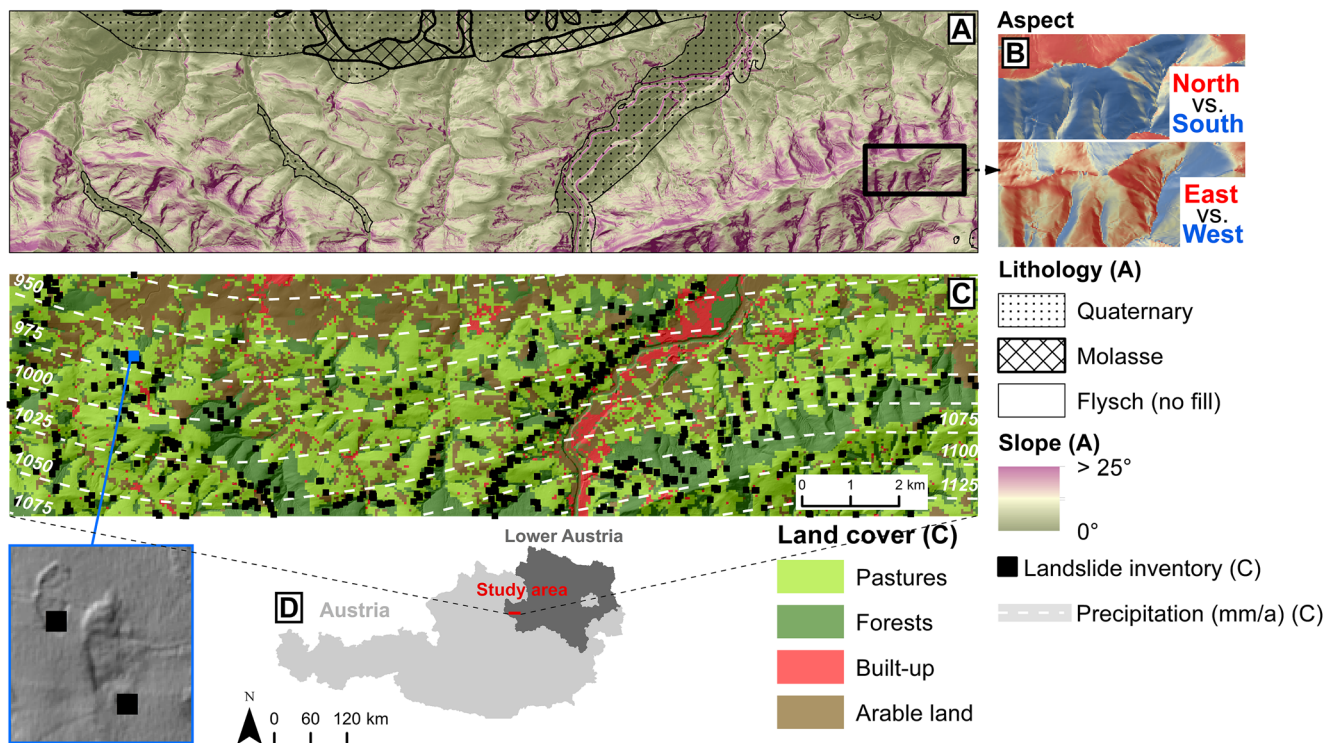


Fig. 1 Location (d) and overview of the study area. Spatial distribution of slope angles (a), lithology (a), slope orientation (b), land cover units (c) and mean annual precipitation (c). The (unmodified) landslide inventory ($n = 591$) is given in c. The shaded relief image (excerpt in blue) shows the geomorphic footprint of characteristic shallow landslides of the area. Corresponding squares relate to the modelling resolution of 10 m × 10 m and depict the landslide scarp mapping location

the South (790 m a.s.l.) (Fig. 1c) (Skoda and Lorenz 2007). The prevalent shallow and small landslides of the area are usually triggered by rainfall and/or snow-melting events. In particular, locally concentrated convective precipitation events in the summer or long-lasting rainfall events between autumn and spring are known to promote shallow landslides of the slide-type movement according to the classification of Cruden and Varnes (1996) and Dikau et al. (1996). Critical landslide-triggering conditions are regularly achieved whenever intensive snow melting coincides with severe precipitation (Schwenk 1992).

The study area is described in more detail in Steger et al. (2016b) while additional information on the prevalent landslide processes and the litho-morphological characteristics of the area and its surroundings can be found in Schwenk (1992), Wessely et al. (2006) and Petschko et al. (2016).

'Real data' and 'synthetic data'

Landslide data (Fig. 1c), topography (i.e. slope, northness, eastness) and lithological information (Fig. 1a) as well as the synthetically generated basic data sets (Fig. 2) were adopted from

previous analyses (cf. Steger et al. 2016b). Within this study, the expression real data relates to real world data from Lower Austria while the expression synthetic data refers to artificially generated data sets.

The available 'real' point-based landslide inventory ($n = 591$) of the study area represents landslide scarps of mainly smaller and shallow landslides of the slide-type movement (cf. excerpt of Fig. 1c) and was mapped by Petschko et al. (2016) for statistical landslide susceptibility modelling, mainly on the basis of high resolution shaded relief images of an airborne laser-scanning digital terrain model (ALS-DTM). Comprehensive information on this landslide data set and its mapping procedure can be found in Petschko et al. (2016).

A number of landslide susceptibility studies advocate to represent each landslide with a single point in order to reduce spatial autocorrelation of observations and to avoid weighting for landslide size (e.g. Van Den Eeckhaut et al. 2006; Atkinson and Massari 2011; Petschko et al. 2014b; Steger et al. 2016b). Several field trips not only verified the high positional accuracy of the present inventory but also emphasized that this data set is likely to be

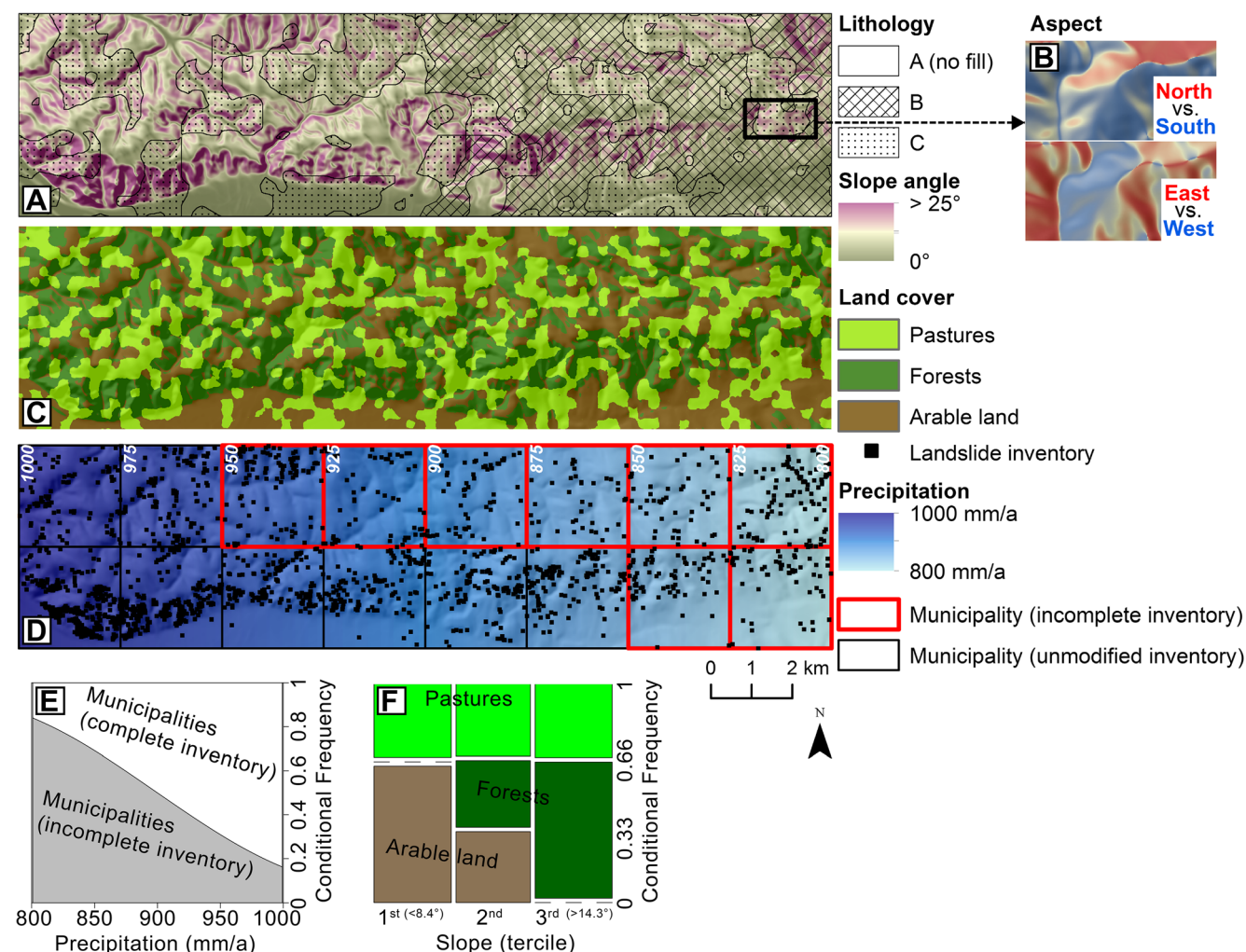


Fig. 2 Spatial representation of the synthetic data sets. Spatial distribution of slope angles (a), lithology (a), slope orientation (b), land cover units (c) and mean annual precipitation (d). The simulated distribution of landslide locations ($n = 2000$) is given in d. The municipalities depicted in d are used to simulate an inventory-based mapping bias related to administrative units. Conditional frequencies in e show that municipalities affected by a simulated municipality-related bias exhibit lower precipitation rates. Areas affected by the simulated forest-related inventory bias (forests) are predominantly located on steeper slopes (f)

biased. This relates in particular to the observation that the extent of anthropogenic interventions (e.g. planation, remediation works) on geomorphic features may vary considerably between different land cover types (Bell et al. 2012). Thus, the inventory is expected to miss an unknown number of landslides located on pastureland or arable land and in close proximity to infrastructure. In contrast, landslides are believed to be overrepresented under forests as a consequence of well-preserved and therefore visually easily detectable geomorphic features (Bell et al. 2012; Petschko et al. 2014a, 2016).

Topographic predictors (Fig. 1a, b; real data) were based on derivatives of a resampled (10 m × 10 m) ALS-DTM whereas lithological information was extracted from a digital geological map (GK200, scale 1:200,000). The reference landslide susceptibility model for the real data was generated using the predictor set 'B', which consists of four variables, namely slope, northness, eastness and lithology. In the following, this set was further enhanced to include the predictors land cover and mean annual precipitation rates to evaluate the effect of including or excluding predictors that are spatially related to a specific inventory-based incompleteness (Fig. 1c; cf. "Methods" section). Land cover was extracted from classifications conducted by Eder et al. (2011) while mean annual precipitation rates were derived from modelling results of Skoda and Lorenz (2007). A municipality layer was used within the applied mixed-effect models to account for a simulated municipality-related inventory bias (cf. "Generalized linear models and generalized linear mixed models" section). Corresponding boundaries were gained from a digital topographic map (ÖK50, scale 1:50,000).

The present synthetic data (Fig. 2) corresponds to data sets described in detail in Steger et al. (2016b). This artificially generated data has already proven valuable to profoundly elaborate the effect of inventory-based positional errors on statistical modelling results as it allowed to define a 'true' and unbiased relation between an assumed perfectly accurate and complete inventory ($n = 2000$; Fig. 2d) and those five environmental factors that were defined to determine landslide susceptibility. Within this study,

synthetic data was specifically utilized to further verify the results obtained by the real data set under controllable conditions.

The predictor set 'B + L' consists of those predictors (i.e. slope, northness, eastness, lithology and land cover) and thus relates to the respective reference landslide susceptibility model of the synthetic data set. Slope angles (Fig. 2a) and both aspect layers (Fig. 2b) were derived from a strongly generalized ALS-DTM while the three equally sized lithological units (Fig. 2a) were spread over the West (unit A), the East (unit B) and across the entire area (unit C). The distribution of the three land cover units was specified to be conditioned on slope angle (Fig. 2c, f).

Within this study, we additionally defined a synthetic mean annual precipitation layer (Fig. 2d) which spatially relates to a simulated municipality-related incompleteness of the inventory (Fig. 2d, e). Corresponding 16 square-shaped municipality boundaries exhibit a spatial extent of 6.25 km² each (Fig. 2d).

Methods

The methodological framework of this study consisted of an artificial introduction of two different types of mapping biases (i.e. forest bias; municipality bias) into the available real landslide inventory and a synthetically generated landslide data set. Logistic regression was applied to model landslide susceptibility separately with each of those differently biased inventories by separately using two different predictor sets (Fig. 3). The first set consisted of predictors which were not specifically related to a simulated inventory incompleteness while the second set contained one environmental factor spatially related to a simulated bias (i.e. land cover for the forest-related bias; precipitation for the municipality-related bias). Furthermore, we tested whether the application of mixed-effects logistic regression models enabled less confounded predictions when modelling with systematically incomplete inventories (cf. "Generalized linear models and generalized linear mixed models" section). Finally, all models were thoroughly evaluated (cf. "Model evaluation" section). All statistical analyses were performed using the statistical software R and its packages 'stats',

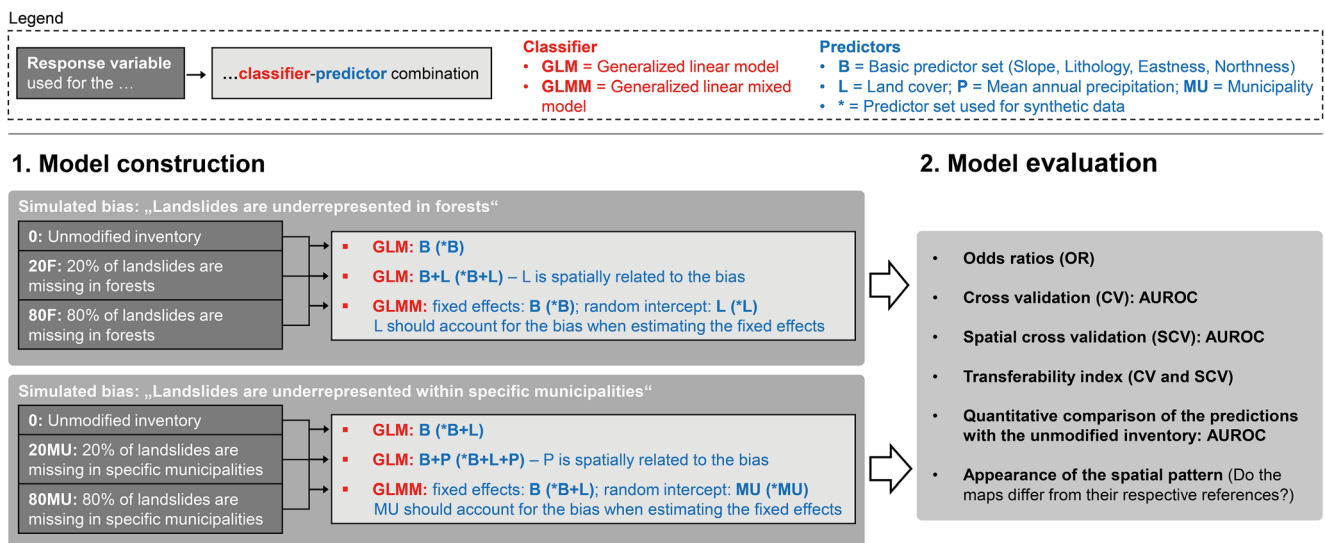


Fig. 3 Methodological framework of this study. After an artificial introduction of two different mapping biases into a landslide inventory (dark grey boxes), landslide susceptibility models were generated separately by considering different classification techniques (red texts) and predictor combinations (blue texts). The subsequent 34 models were evaluated with multiple techniques (model evaluation). Note that abbreviations (e.g. 'GLM: B + L') are further used within upcoming figures

'lme4' and 'sperrorest' (Brenning 2012a; Bates et al. 2014; R Core Team 2014).

Artificial introduction of inventory-based biases

The first inventory mimics a systematic forest-related bias which may arise when landslides are mapped by solely interpreting aerial photographs (Brardinoni et al. 2003) or when landslides in forest areas are underreported (Steger et al. 2016a). A forest-related incompleteness was artificially introduced in this study by randomly deleting 20 and 80% of landslides within areas classified as forests (Fig. 4a).

A spatially varying bias may also arise from merging different landslide data sets from adjacent areas (Van den Eeckhaut et al. 2012) especially if different mapping procedures were used, if data sets were compiled for different purposes or by individuals with different levels of expertise, or if underreporting varies between jurisdictions (Ardizzone et al. 2002). In order to mimic this possible source of bias, a second bias was introduced in this study by gradually removing 20 and 80% of landslides within specific municipalities to simulate a bias related to the administrative units (Fig. 4c).

Generalized linear models and generalized linear mixed models

Binary logistic regression is commonly applied to predict landslide susceptibility at a regional scale. This classifier is based on a generalized linear model (GLM) with a logistic link function that enables modelling the (fixed) effect (cf. fixed part of Eq. (1)) of each predictor or predictor class on the response (landslide presence/absence) (Atkinson et al. 1998; Brenning 2005; Van Den Eeckhaut et al. 2010; Regmi et al. 2014; Felicísimo et al. 2013; Budimir et al. 2015). Binary logistic regression is further referred to as GLM.

A confounder is a variable which is associated with both, the response variable (i.e. landslide inventory) and another variable

(e.g. slope) (Brenning 2012b; Szklo and Nieto 2014). In the context of systematically incomplete inventories, confounding may become particularly problematic whenever an inventory-based incompleteness is directly related to a specific variable (e.g. land cover class, municipality) that acts as a confounder for other predictors (e.g. slope, lithological units). For instance, in the likely case that forested areas are more frequently located on steeper terrain (Rickli et al. 2002; Steger et al. 2016a; cf. Fig. 4b), a forest-related incompleteness of the inventory may lead to the tendency that landslides are underrepresented not only in wooded areas but also on steeper slopes. Thus, a potential exclusion of the bias-describing predictor land cover may lead to systematically confounded modelling results (e.g. in the form of a biased slope coefficient) (Brenning 2012b; Steger et al. 2016a). However, earlier findings also provided evidence that an inclusion of such a bias-describing predictor may as well be related to misleading modelling results because of a subsequent direct bias propagation via the included variable (Steger et al. 2016a).

We aimed to tackle the problems of direct bias propagation and confounded relationships by proposing a novel statistical approach (i.e. mixed-effects models) to improve landslide susceptibility models generated with systematically incomplete inventories. From our knowledge, all statistical landslide susceptibility models generated up to now belong to the group of fixed-effects models, where the main interest lies in the specific influence of each predictor and predictor level on the response (Bolker et al. 2009). Mixed-effects models additionally allow us to include random effects and have proven to be useful to analyze nested (e.g. hierarchically grouped) or correlated (spatial or temporal) data in the fields of medicine, economy, social science and ecology (Zuur et al. 2009). An inclusion of random terms may also be valuable when a specific categorically scaled predictor is considered to be a 'nuisance' parameter which is not of direct interest, but should be accounted for when estimating the fixed-effects coefficients (Bolker et al. 2009). Within this study, mixed-effects logistic

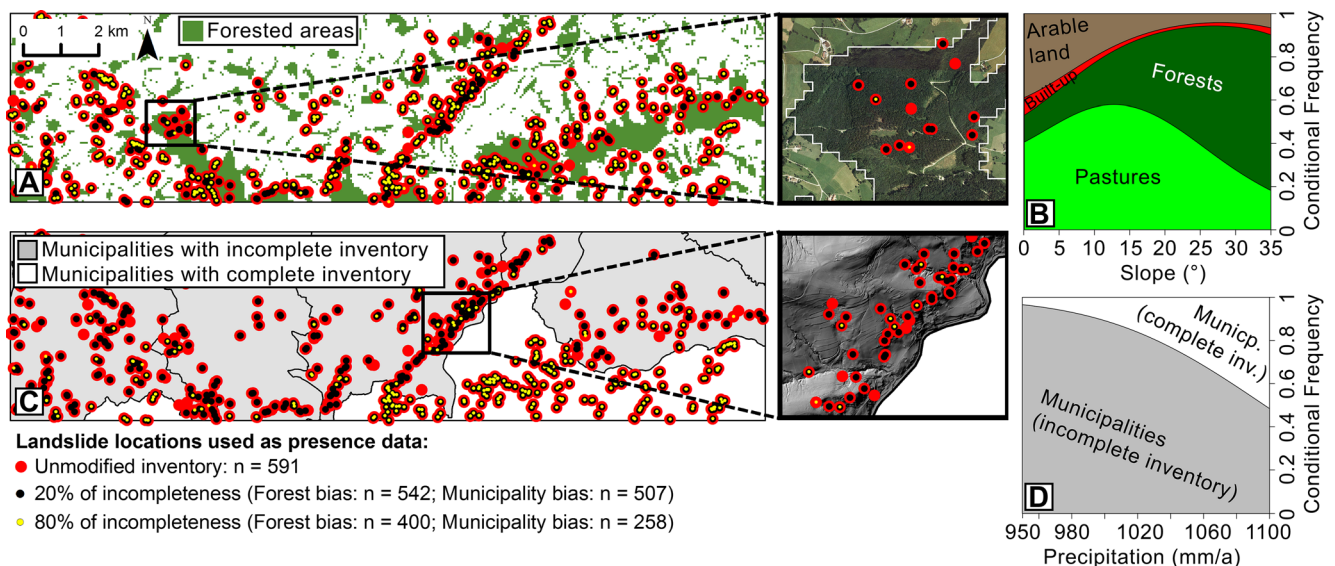


Fig. 4 Unmodified landslide inventory (red dots in a and c) and simulated inventory-based incompleteness within forested areas (black and yellow dots in a) and within specific municipalities (black and yellow dots in c) for the real data set. The conditional frequency plots (b, d) show observed spatial interrelations between variables affected by a bias and environmental factors: Forests are more likely located on steeper slopes (b) and municipalities affected by a bias (grey area) generally show lower precipitation rates (d)

regression models were applied to estimate a binary outcome (i.e. landslide presence/absence). More specifically, we fitted a generalized linear mixed model (GLMM) while specifying a bias-describing variable as random intercept and the other predictors as fixed effects:

$$\text{logit}(P(Y_{ij} = 1)) = \underbrace{\beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}}_{\text{fixed effects}} + \underbrace{\gamma_j}_{\text{random intercept}} \quad (1)$$

where β_0 relates to the intercept, $\beta_1 \dots \beta_p$ to the regression coefficients of associated predictors $X_1 \dots X_p$ and γ to the random intercept. γ is assumed to have a prior distribution which is normally distributed with mean zero and variance σ^2 (Bolker et al. 2009; Zuur et al. 2009; Bates et al. 2014). The idea behind this procedure was to separate (i.e. isolate) the variation which relates to inventory-based bias from the effects which are assumed to be less influenced by those biases. Thus, the random intercept term was specifically included to represent (i.e. account for) a bias which is directly related to a categorically scaled variable when estimating the coefficients of the fixed effects predictor variables. The variable land cover was included as a random intercept to represent variations originating from a forest-related bias while a municipality layer was introduced to account for the bias directly related to the respective administrative boundaries. Finally, only the fixed effects were used to predict landslide susceptibility. Thus, all presented GLMM predictions as well as related predictive performance estimates were based on a random intercept of 0 (i.e. fixed effects alone), averaging out the random effects.

Model evaluation

The odds ratio (OR) represents a measure of association and is regularly used to compare modelled relations within logistic regression models (Hosmer and Lemeshow 2000). The modelled relationships between the response variable and the predictors were evaluated by estimating ORs for 'meaningful' increments of each predictor (Brenning 2012b; Brenning et al. 2015; Steger et al. 2016b). For instance, ORs estimated for the predictor slope displayed differences in the chance that a 10° steeper slope is affected by future landsliding (e.g. 1 = equal chances, >1 steeper slopes are more likely affected). This accounts for potential confounding effects of other predictors included in the model. In analogy, ORs obtained for single land cover classes depict how much more likely (or unlikely) a specific land cover type (e.g. forest) is affected by landsliding compared to a reference class (e.g. pastures). This consideration already accounts for the effect that forests are more frequently located on steeper slopes (Steger et al. 2016a).

Comparisons of modelled relationships and landslide susceptibility maps with their references (i.e. models assumed to be less affected by a bias) provided indications on how the respective inventory-based errors were reflected by the final results. A substantial deviation of ORs and predicted susceptibility patterns would be interpreted as evidence that the inventory mapping error was propagated into the final modelling results. Quantile classification of the final maps was conducted to ease a visual comparison of susceptibility patterns (Hussin et al. 2016).

The predictive performance of all models was assessed by estimating the AUROC (0.5 = random model; 1 = perfect discrimination between landslides and non-landslides) by means of a

repeated non-spatial (cross-validation; CV) and spatial (spatial cross-validation; SCV) partitioning of training and test samples (Brenning 2012a; Petschko et al. 2014b; Goetz et al. 2015). CV and SCV was based on a 50-repeated 10-fold validation for each model.

The transferability index, which is also based on a repeated estimation of predictive performances, was adopted from Petschko et al. (2014b) and provided information on the non-spatial (CV) and spatial (SCV) transferability of the modelling results. This metric mainly reflects the interquartile range of computed AUROCs while additionally accounting for a variability of sample sizes in the respective test data set. A low transferability index points out that the obtained predictive performances of a model were relatively similar for different partitions. Thus, obtained low index values indicate robust modelled relationships and a subsequent high non-spatial (CV) and spatial (SCV) transferability of modelling results (Petschko et al. 2014b).

The predicted susceptibility score was further compared with the unmodified response variable using the AUROC. This measure provided quantitative information on how well the models were able to predict landslide observations related to less biased (real data) and unbiased (synthetic data) response variables. Note that this measure corresponded to the goodness of model fit whenever the respective data sets related to the unmodified inventories (0% of simulated incompleteness).

Results

The impact of the land cover-related inventory bias

ORs obtained for the GLMs generated without the land cover variable, but with highly biased inventories, provided evidence of confounded relationships between landslide inventories and those predictors that were correlated with these bias-describing variables (e.g. forests are more likely located on steeper slopes; Fig. 4b). For instance, a substantial (80%) forest-related underrepresentation of landslides led to a weaker modelled dependence of landslide occurrence on slope angle ('GLM: B' in Fig. 5a, e). ORs (reported with 95% confidence intervals in square brackets) of the study area's reference model ('R' in Fig. 5a) showed that the chances of a 10° steeper slope to be affected by landsliding were 7.3 [5.5/9.8] times higher compared to their 10° flatter counterparts while the respective model generated with the highly biased inventory ('GLM: B', 80% in Fig. 5a) exposed considerably lower ORs 5 [3.7/7]. ORs obtained from the synthetic data sets exposed that the modelled associations between slope angles and landslide occurrence differed between the reference model ('R' in Fig. 5e) and the models generated without land cover ('GLM: B'). This tendency was further intensified with an increasing portion of inventory-based incompleteness on forested areas ('GLM: B' in Fig. 5e). The respective ORs of the predictor slope showed a decrease from 5.4 [4.8/6.1] to 4 [3.5/4.5].

ORs of models generated with land cover as a predictor ('GLM: B + L') and differently biased inventories (0, 20, 80%) provided quantitative evidence that land cover partly accounted for a variability originating from the simulated forest-related bias, when estimating the fixed-effects coefficients of the other predictors. Thus, ORs estimated for the predictor slope were more similar across all simulated biases (cf. dashed line with black dots in Fig. 5a, e). For example, GLMs fitted with the highly biased inventory (80%) showed an OR deviation of the predictor slope from

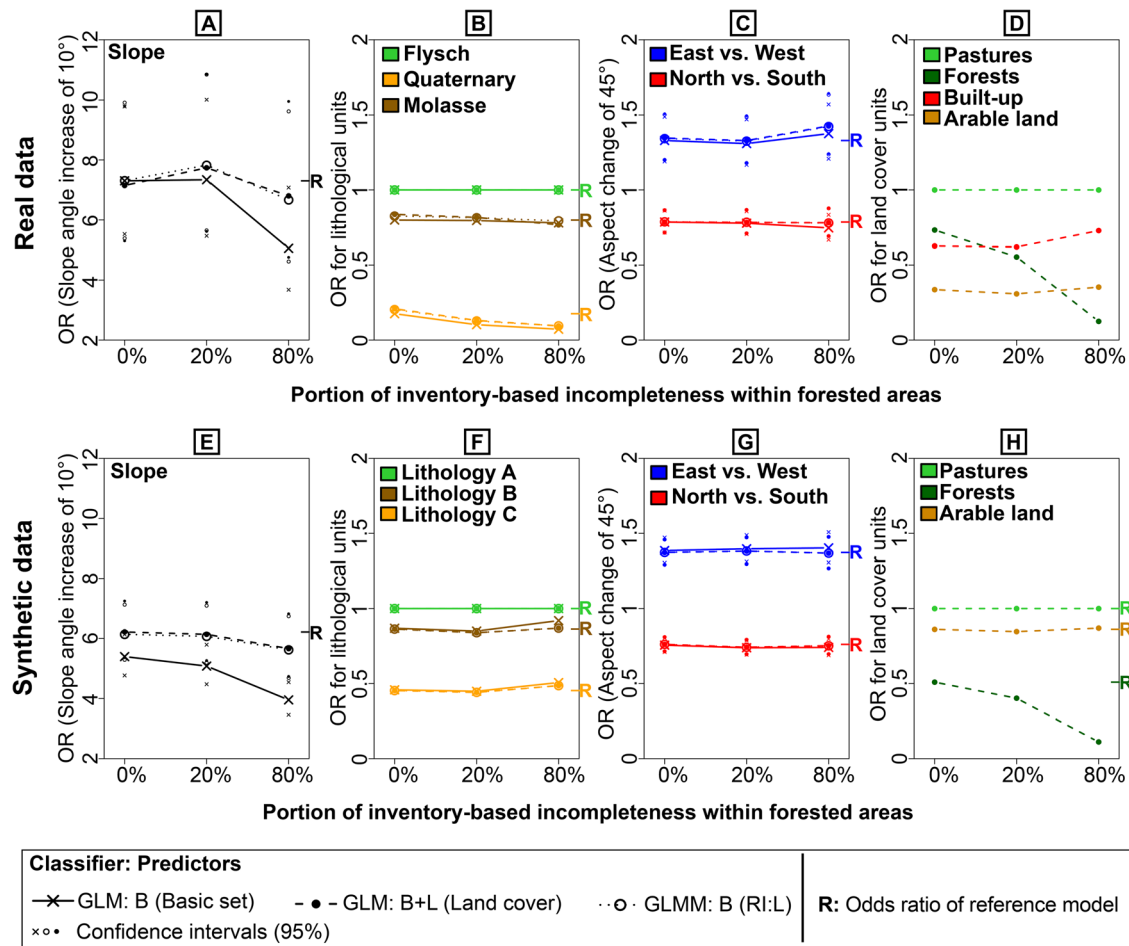


Fig. 5 Modelled relationships expressed as odds ratios (OR) for models generated with differently incomplete inventories on forested areas (0, 20, 80%) for the real data set (a, b, c, d) and the synthetic data (e, f, g, h). GLMs generated with a highly biased inventory (80%) and without land cover as a predictor ('GLM: B') provide evidence of confounded slope coefficients (a, e). GLMs generated with land cover as a predictors ('GLM: B + L') indicate a direct bias propagation into the final model via the predictor land cover (d, h). GLMMs avoided this bias and simultaneously reduced confounding effects

the respective reference models of 0.9 (real data) and 0.5 (synthetic data) whenever the model was generated with land cover as a predictor and 2.3 (real data) and 1.4 (synthetic data) in the case the model was produced without land cover. However, modelled relationships obtained for the predictor land cover (Fig. 5d, h) provided further quantitative evidence that a simulated inventory-based bias may be directly propagated into the final models when a specific predictor (in this case land cover) relates to a systematic incompleteness of the inventory (Steger et al. 2016a). Consequently, GLMs predicted constantly decreasing and finally very low chances of forests to be affected by future slope movements (i.e. OR drop to 0.1 in Fig. 5d, h), ultimately due to the high number of missing landslides in forested areas. The resulting landslide susceptibility maps (Fig. 6f, o) directly reflected this bias at forest locations (Fig. 6s, t) by showing considerably lower susceptibility values in comparison to their references (Fig. 6a, m).

In general, ORs obtained for the GLMMs provided evidence that these models accounted for inventory bias while additionally ensuring that the bias was not directly propagated into the final models via the land cover predictor. The influence of forest-related variation was accounted for by the random intercept for land

cover. Thus, GLMMs and GLMs generated with land cover as a predictor ('GLM: B + L') showed similar and more stable modelled associations, especially for the predictor slope, than GLMs without land cover (Fig. 5). The previously mentioned direct bias propagation via the predictor land cover was successfully avoided, because the respective predictions were based on the fixed effects alone. Therefore, whenever the respective susceptibility maps were generated with the highly biased inventory ('80F' in Fig. 6), spatial patterns of GLMM-based maps (Fig. 6i, r) appeared relatively similar to the reference maps (Fig. 6a, m).

Validation results continuously showed AUROC values of >0.8 (Fig. 7). A comparison of predictive performances provided evidence that GLMs generated without land cover (Fig. 7a, d) and GLMMs (Fig. 7c, f) performed worst when the respective models were generated with a strongly biased inventory (80%). A contrasting trend and the highest predictive performances were observed for the models that were previously identified as being highly biased ('80%' in Fig. 7b, e). However, further comparisons exposed their poorest performance in predicting the original landslide locations (grey line in Fig. 7). This discrepancy was interpreted as quantitative evidence for overoptimistic predictive

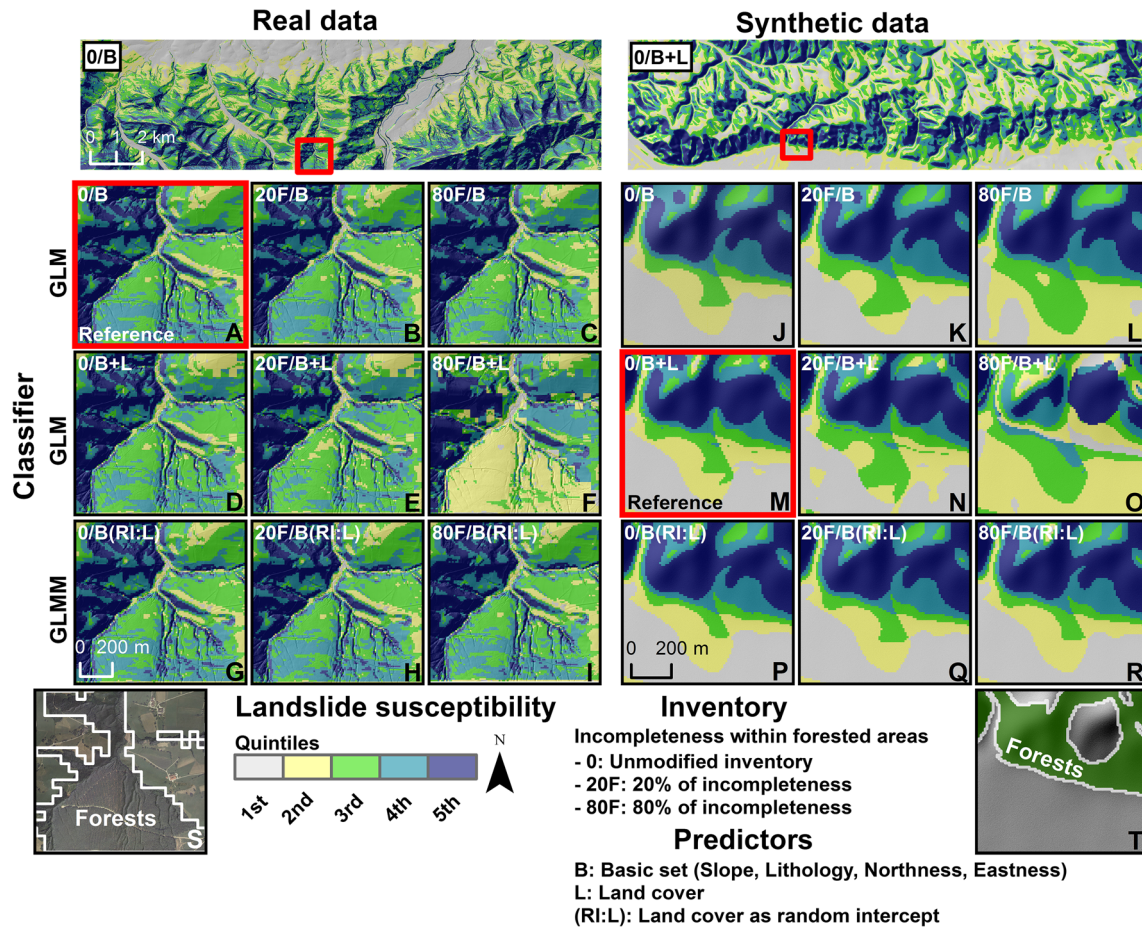


Fig. 6 Excerpts of landslide susceptibility maps generated with differently biased inventories on forested areas (0, 20, 80%) by applying different classifier-predictor combinations for the real data set (a, b, c, d, e, f, g, h, i) and the synthetic data (j, k, l, m, n, o, p, q, r). The maps marked as 'Reference' are considered to be only slightly influenced (a) or unaffected (m) by an inventory-based bias. All other maps are interpreted relatively to these reference maps. Note the substantial differing appearance of susceptibilities at forested areas (s, t) for all maps generated with a highly biased inventory and land cover as a predictor (f, o) and similarly appearing GLMM-based maps

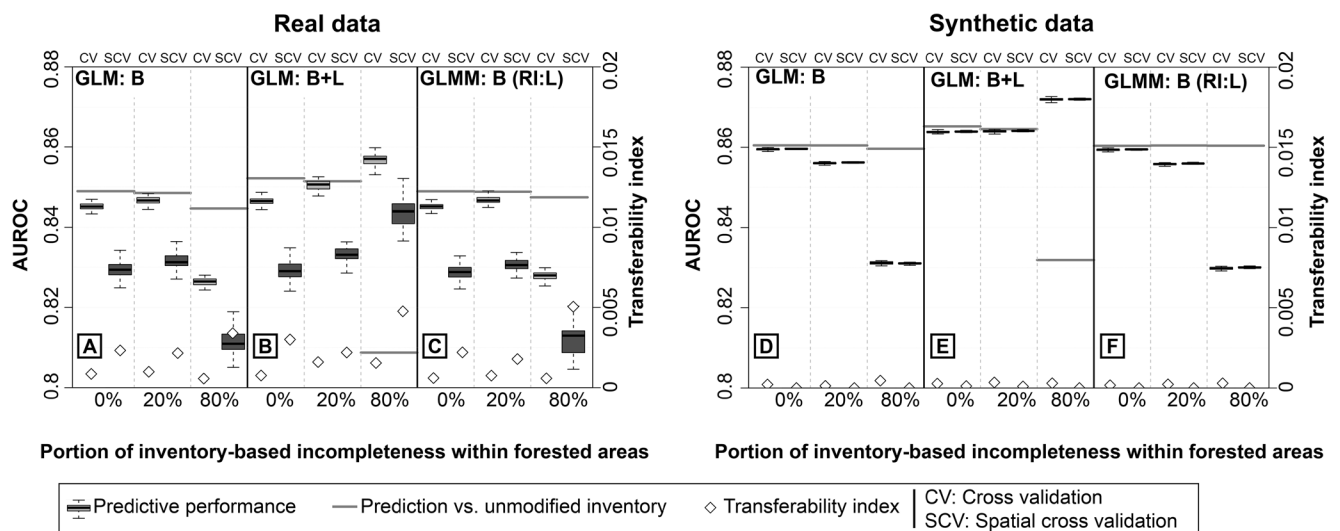


Fig. 7 Validation results and transferability indexes obtained for models generated with differently complete inventories in forested areas for the real data set (a, b, c) and the synthetic data (d, e, f). Boxplots refer to predictive performances obtained by cross-validation (CV) and spatial cross-validation (SCV). The grey line shows a comparison of model predictions (AUROC) with the data set that relates to the unmodified inventories (= model fit for all 0%-models). Diamonds relate to the second y-axis and show non-spatial (CV) and spatial (SCV) transferabilities of the modelling results

performance estimates for GLMs generated with land cover as a predictor ('GLM: B + L').

Generally, higher transferability indexes were obtained for the models generated for the real data set. The observed low variation in predictive performances (i.e. low transferability index in Fig. 7) obtained for the models generated with synthetic data indicated a high non-spatial (CV) and spatial (SCV) transferability of the modelling results. Lowest spatial transferabilities (highest index values for SCV in Fig. 7) were assigned to the real data models based on highly incomplete inventories.

The impact of the municipality-related inventory bias

The main insights obtained by simulating a municipality-related inventory bias were, with exceptions, similar to the trends detected by mimicking a forest-related incompleteness. GLMs based on the most incomplete inventory (80%) and without a predictor that was spatially related to a simulated bias (in this case precipitation) generally showed the most distinct deviations of modelled relationships from the reference models (cf. compare the crosses with 'R' in Fig. 8). This in turn provided evidence for the presence of confounded relationships. The resulting biases can also be traced back by visually comparing the respective landslide susceptibility maps (Fig. 9c, l) with their references (Fig. 9a, j).

Generally smaller distortions (i.e. changes in ORs) were observed for the predictor slope in the case precipitation, which is spatially related to the bias (Figs. 2e and 4d), was included as a predictor (cf. compare black dots with 'R' in Fig. 8a, e). However, a direct bias propagation within those models was exposed when interpreting ORs for a precipitation increase of 50 mm (Fig. 8d, i).

For instance, the respective ORs obtained for models generated with the unmodified inventory provided evidence of a non-existing and slightly negative association between the respective landslides and the mean annual precipitation rates (Fig. 8d, 0.92; Fig. 8i, 0.97). However, an apparent strong and positive relationship emerged in models fitted to an 80% incomplete inventory (Fig. 8d, 1.86; Fig. 8i, 1.47). This spurious relationship was well visible in the resulting susceptibility maps (compare Fig. 9f, o to Fig. 9a, j).

GLMMs were again able to account for some variation related to a simulated inventory-based incompleteness while additionally avoiding a direct bias propagation (i.e. since a bias-describing predictor was not used to predict susceptibility). Therefore, susceptibility maps obtained from GLMMs with the highly biased inventory (Fig. 9i, r) appeared most similar to their references (Fig. 9a, j).

However, none of the models were able to avoid a substantial decrease in estimated OR within the lithological unit Molasse (Fig. 8b). This relatively small area (3% of total area) was observed to be entirely covered by the municipalities affected by a bias (compare Fig. 1a with Fig. 4c) while additionally being represented by a very small sample size ($n = 21$). The OR decrease was accompanied by a decrease in the ratio between landslide presences and absences. We observed a considerable decrease in this ratio from an initial 17% (3 landslides) to 6% (1 landslide) and 0% (no landslide) with increasing incompleteness of the inventory. Consequently, all landslide susceptibility maps generated with the highly biased inventory erroneously indicated that the Molasse Zone is stable, regardless of the other environmental conditions.

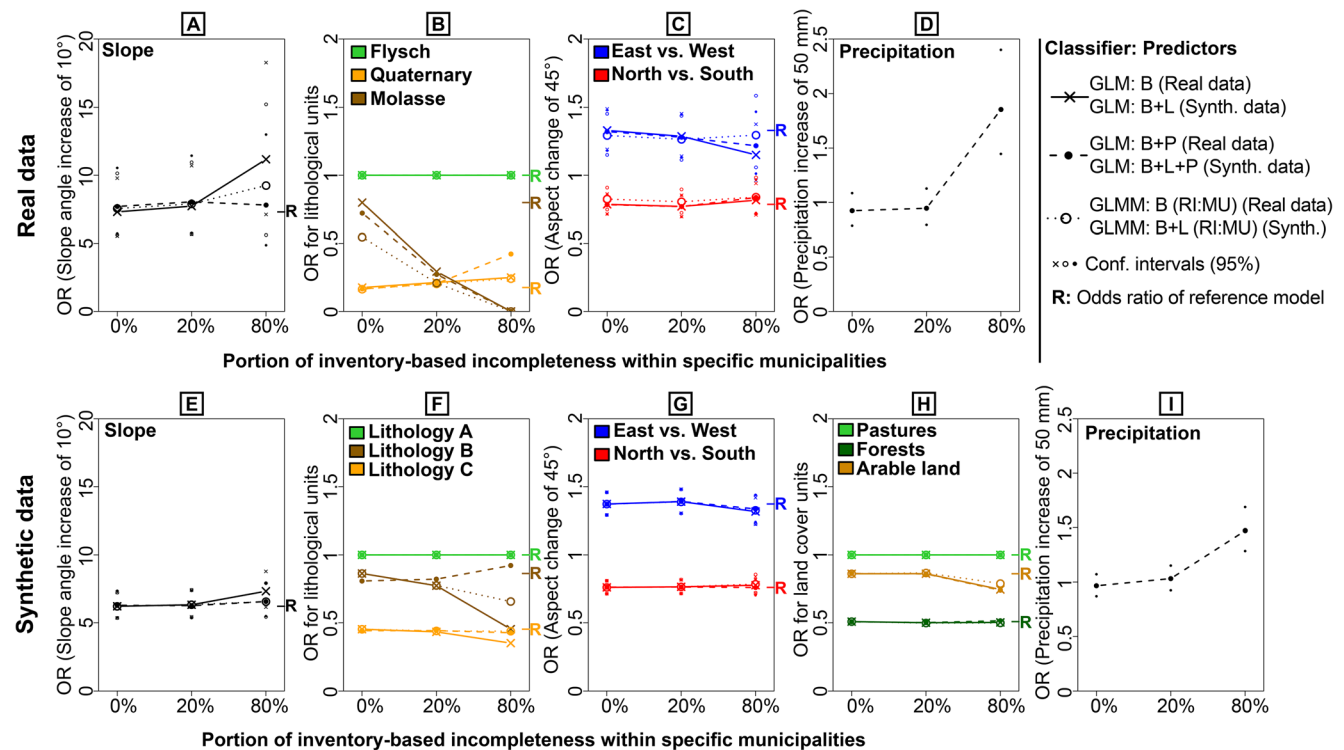


Fig. 8 Modelled relationships expressed as odds ratios (OR) for models generated with differently complete inventories (0, 20, 80%) in specific municipalities (cf. Figs. 2d and 4c) for the real data set (a, b, c, d) and the synthetic data (e, f, g, h, i). GLMs generated with a highly biased inventory (80%) and without precipitation ('P') as a predictor indicated high deviations in modelled relationships from the reference models ('R'). ORs of the predictor precipitation (d, i) provide evidence that an actually non-existing relation (OR near 1 at 0%) was turned into an apparent positive association. GLMMs avoided this bias while simultaneously reducing confounding effects

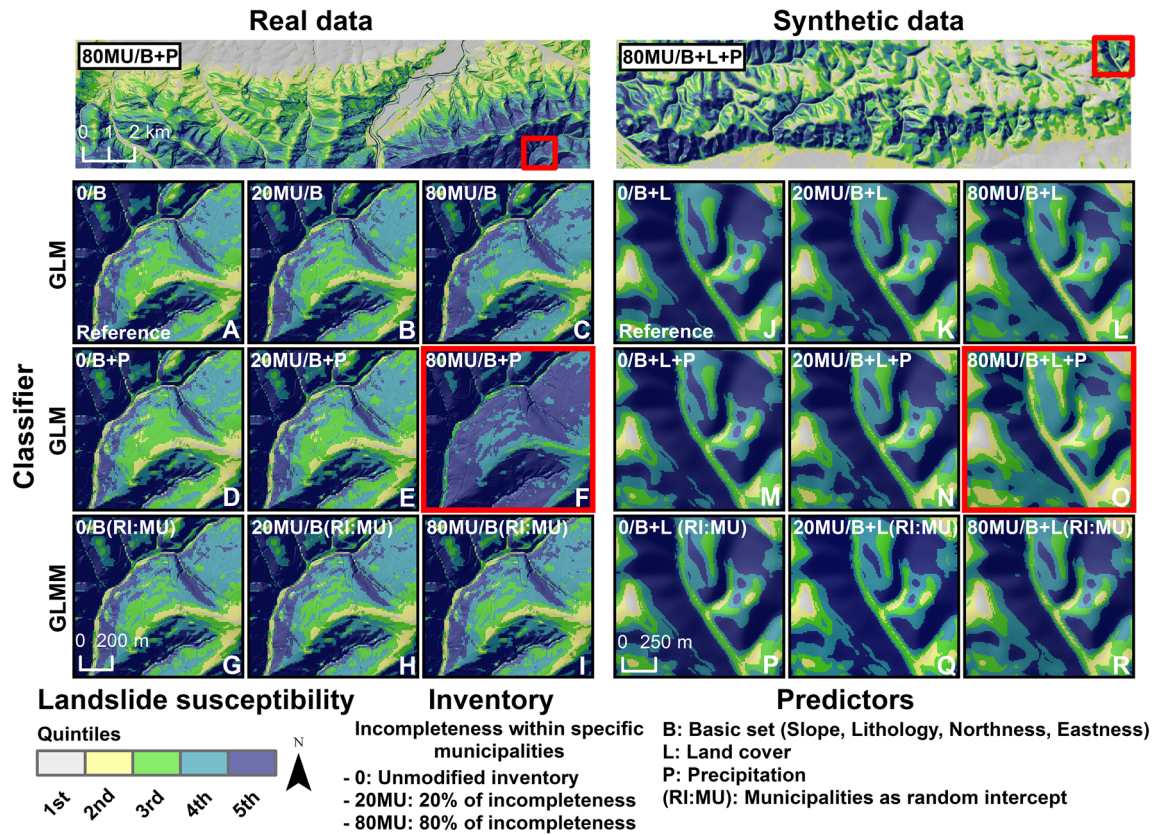


Fig. 9 Excerpts of landslide susceptibility maps generated with differently biased inventories (0, 20, 80%) on specific municipalities for an area with high precipitation rates (real data, a, b, c, d, e, f, g, h, i) and low precipitation rates (synthetic data, j, k, l, m, n, o, p, q, r) by applying different classifier-predictor combinations. The maps marked as 'Reference' are considered to be only slightly influenced (a) or unaffected (j) by an inventory-based bias. All other maps are interpreted relatively to these maps. Note the differences between the References and the respective maps generated with a highly biased inventory and precipitation (f, o). GLMM-based maps (i, r) appeared most similar to the reference maps (a, j) in the case only highly biased inventories were available ('80MU')

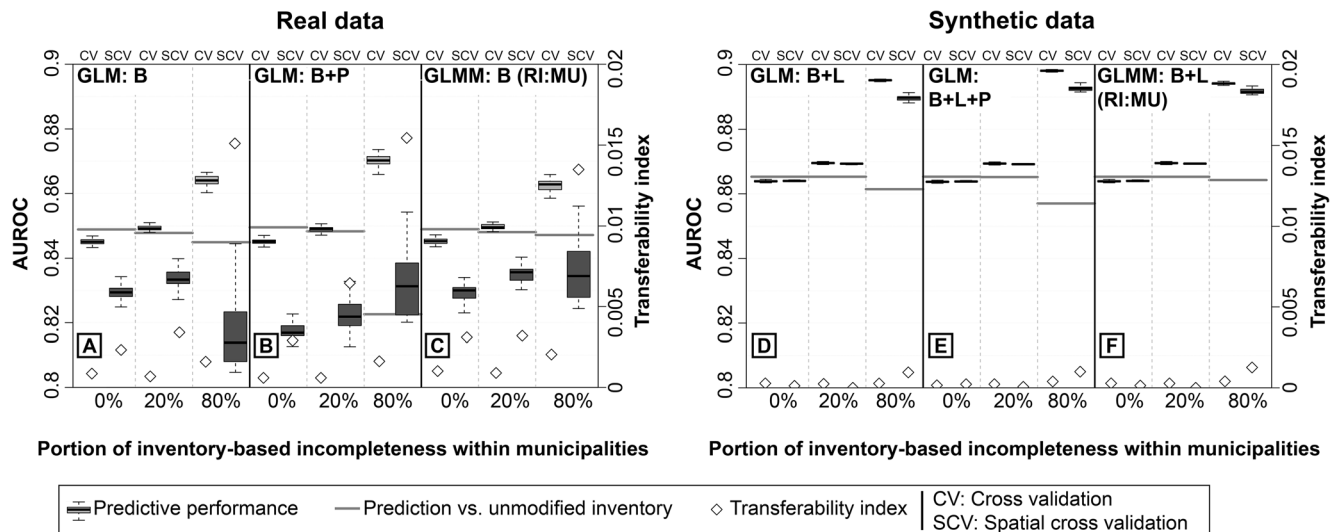


Fig. 10 Validation results and transferability indexes for models generated with differently biased inventories (0, 20, 80%) on specific municipalities for the real data set (a, b, c) and the synthetic data (d, e, f). Boxplots refer to predictive performances obtained by cross-validation (CV) and spatial cross-validation (SCV). The transferability index (second y-axis) exposed a lower internal spatial transferability of modelling results (i.e. high SCV based index) as a result of the simulated spatial inventory-based bias. The grey line depicts the comparison of model predictions with the original landslide position (= model fit for all 0%-models). In the case only a highly biased inventory was available (cf. all 80%-models), GLMMs (c, f) performed best to predict the unmodified inventory while the apparently best performing GLMs generated with precipitation (cf. CV-based predictive performance in b and e) performed worst

Model validation (Fig. 10) consistently produced AUROC values higher than 0.8, which would normally be considered to reflect a good (Fressard et al. 2014) or excellent (Conoscenti et al. 2016) predictive performance of a landslide susceptibility model. Comparing these estimates, we observed that the apparent predictive performance obtained by CV increased with an increasing bias of the inventory. Models that included a bias-describing predictor exposed highest CV-based AUROCs (80% in Fig. 10b, e), but the lowest ability to predict the original landslide positions (grey line in Fig. 10b, e). In this regard, GLMMs generated with the highly biased inventory performed slightly better, compared to the other models generated with an evenly biased data set.

SCV revealed remarkably higher variations of AUROC values in the case the underlying models were generated with the highly biased inventory (80%). The lower spatial transferability of those models is reflected by high transferability indexes and can visually be examined by comparing the respective boxplot sizes (Fig. 10). Further indications of spatially inconsistent modelling results were obtained by comparing CV-based AUROCs with SCV-based AUROCs of identical models. Predictive performances of the synthetic models, which in other cases were similar (compare CV and SCV in Fig. 7d–f), discernibly dropped when spatially estimated for the highly biased data set ('80%' in Fig. 10d–f). AUROCs of identical models generated with real data were also remarkably lower when assessed in a spatial context (Fig. 10a–c, but also Fig. 7a–c).

Discussion

Our study confirmed and provided further quantitative evidence of the critical importance of landslide inventory completeness for the quality and validity of statistical landslide susceptibility assessments (Ardizzone et al. 2002; Galli et al. 2008; Harp et al. 2011; Fressard et al. 2014; Steger et al. 2016a, b). However, our results also emphasized that linkages between the completeness of a landslide inventory and modelled relationships, validation results and the spatial appearance of the final maps are multi-faceted.

In particular, the inventory's degree of completeness is only one of the several aspects that determine how and to what extent an inventory bias may propagate into the results. Apart from the number of the respective observations within specific predictor classes (cf. "Minor inventory biases locally affect modelling results" section) and the extent of spatial agreement between predictors and areas affected by an inventory bias (cf. "The influence of bias-describing predictors" section), the selected modelling approaches (cf. "Confounding factors and the usefulness of mixed-effects models" section) were also observed to control the influence of systematically incomplete inventories on landslide susceptibility models. Whether or not a modeller is able to detect such discrepancies may in turn depend on the choice of the cross-validation technique used to evaluate predictions (cf. "The influence of bias-describing predictors" and "The need for differentiated model evaluations" section). Based on our findings, we finally propose a four-step procedure to deal with systematically incomplete inventories in the context of statistical landslide susceptibility modelling (cf. "Practical recommendations" section).

Minor inventory biases locally affect modelling results

It was remarkable that landslide susceptibility models generated with a 20% systematically incomplete inventory did not

substantially differ from the respective reference models. We observed similar modelled relationships, comparable predictive performances and, with local exceptions, visually similar maps, especially in situations where a bias-describing predictor was not included. Due to model generalization and model uncertainty, systematic incompleteness of this order of magnitude therefore appeared to be too small to be detectable statistically in landslide susceptibility modelling. Since strongly generalizing classifiers are expected to offer some degree of protection against a direct inventory-based error propagation, we advise to avoid classification techniques that are highly flexible (e.g. machine learning) in the case only incomplete or inaccurate inventories are available (Steger et al. 2016b).

However, the observed substantial OR changes of the relatively small Molasse Zone (Fig. 8b) indicated that a minor inventory incompleteness may locally strongly influence spatial predictions, particularly when a class of a categorically scaled predictor exhibits a small number of observations. The observed high level of uncertainty around the respective coefficient of the reference model (95% OR-confidence, 0.2 to 2.6) additionally indicated its high sensitivity to minor changes in the sample. From this observation, we infer that an inclusion of more detailed thematic information (e.g. lithology, land cover, soil types) does not necessarily lead to improved landslide susceptibility assessments, since a small number of observations within specific classes may result in very uncertain estimates (Heckmann et al. 2014) and a consequent high sensitivity to minor inventory-based biases. When validating the models, it should always be noted that overall, strongly summarizing measures of diagnostic accuracy, like the AUROC (Swets 1988), are not designed and able to detect such local distortions.

In statistical landslide susceptibility modelling, the problem of small sample sizes is regularly counteracted by artificially increasing the number of observations (Poli and Sterlacchini 2007; Fressard et al. 2014). However, sampling multiple points per landslide observation may often not be suitable due to an increasing spatial autocorrelation of cases (Van Den Eeckhaut et al. 2006; Atkinson and Massari 2011), a subsequent overoptimistic and misleading confidence of model parameters (i.e. confidence of coefficients) and a potentially undesired weighting for size (i.e. not providing equal treatment of small and large landslides). Furthermore, feasibility might become an issue in the case that computational demanding state-of-the-art algorithms (e.g. CV-based model parameterization of machine learning techniques, k-fold spatial cross-validation, permutation-based variable importance assessment) are applied (Brenning 2012a).

The influence of bias-describing predictors

The results provided quantitative evidence to the suspicion that distorted relationships and misleading predictive performances may follow whenever a specific predictor systematically relates to a substantial bias of an inventory (Steger et al. 2016a). Specifically, substantially incomplete inventories in forest areas led to spurious modelled relationships (decreasing ORs of forests in Fig. 5d, h) and susceptibility maps (low susceptibility in forests in Fig. 6f, o) as soon as land cover was introduced as a predictor.

The findings additionally indicated that a spatially strongly varying completeness of an inventory may, coincidentally, be spatially related to certain environmental conditions (e.g. municipalities affected by a bias experience lower precipitation rates; cf. Fig. 4d). An inclusion of

such a bias-describing predictor led to a direct bias propagation into the modelling results, which was reflected by the trend that an initially geomorphically implausible negative association between landslide occurrence and precipitation developed into an apparent distinct and influential positive modelled relationship (ORs in Fig. 8d, i). Ultimately, this led to exaggerated landslide susceptibilities at locations with high precipitation (e.g. Fig. 9f) and an underestimation of landslide susceptibility in areas represented by relatively low precipitation (e.g. Fig. 9g). Thus, this study further underlines difficulties which may arise when determining landslide driving factors with statistical models (Vorpahl et al. 2012; Brenning et al. 2015) by providing quantitative indications that inventory errors may influence the weighting of predictors and consequently also the appearance of the final landslide susceptibility maps (Ardizzone et al. 2002; Fressard et al. 2014). Thus, we agree that landslide susceptibility models, particularly generated with biased input data, may not be useful to derive a causal association between environmental conditions and landslide occurrence (Donati and Turrini 2002; Felicísimo et al. 2013), especially since we detected that these biased models might perform better from a purely quantitative perspective.

It was remarkable that an inclusion of previously discussed spurious relations (via a bias-describing predictor) led to increased predictive performances (CV in Fig. 7b, e and Fig. 10b, e). The observation that highest AUROCs may as well be obtained for the geomorphically most implausible maps is in line with previous results (Steger et al. 2016a). We suggest that distorted performance estimates for highly biased models can be expected when firstly, the respective training and test sets are similarly affected by a bias (e.g. both inventory subsets are incomplete in forests). This tendency may be common when applying conventional partitioning techniques (i.e. holdout validation, CV) due to the systematic nature of many biases. Secondly, the respective inventory incompleteness favours an enhanced ability of the bias-describing predictor to distinguish landslide observations from non-landslide observations.

A quantitative comparison of model predictions with the location of the less biased (real data) and unbiased (synthetic data) inventories provided further evidence that the obtained prediction performances can indeed be referred to as overoptimistic, in particular whenever a bias-describing predictor is included into a model. This is another reason why we believe that modellers should not aim to solely improve statistical performance measures like the AUROC by iteratively opting for modifications that increase the respective prediction skills. We finally argue that a process-related interpretation of plausible appearing modelled relationships may be misleading due to possible confounding with inventory errors (Malamud et al. 2004; Guzzetti et al. 2012).

Confounding factors and the usefulness of mixed-effects models

The topography of an area is known to be related to its lithology (Huggett 2007) while topographic variables (e.g. slope, exposition) co-determine land use and thus land cover (Rickli et al. 2002). This example illustrates that environmental factors, which control landslide occurrence, are inevitably interrelated, and thus, some sort of confounding may be regularly present within many statistical landslide susceptibility models (Brenning 2012b). Confounding is the reason why we believe that simply ignoring (i.e. excluding) bias-describing predictors may not be straightforward. We showed that a model fitted on highly biased data, but without a predictor

accounting for the variability originating from such biases, adjusted modelled relationships of predictors that correlated with this incompleteness. For example, a substantial forest-related incompleteness of the inventory led to a decreased sensitivity of the models to slope angle (Fig. 5a) since forests were observed (real data) and simulated (synthetic data) to be more likely located on steeper slopes. In analogy, a considerable municipality-related bias led to decreased modelled susceptibilities within those lithological units (i.e. Molasse and unit B) located inside municipalities affected by an incompleteness of an inventory.

The models generated with a bias-describing predictor (i.e. land cover for the forest-related bias, precipitation for the municipality-related bias) showed the ability to reduce the effects of such confounding, but simultaneously produced highly distorted predictions. The main disadvantages of both previously mentioned procedures were avoided by applying GLMMs with a bias-describing variable introduced as a random intercept. GLMMs proved useful to separate some effects (i.e. variability) related to the bias from the effects assumed to be primarily related to landslide susceptibility. The resulting predictions were less confounded (i.e. similar to models generated with a bias-describing predictor) and not directly affected by the inventory-based incompleteness (i.e. similar to models generated without a bias-describing predictor). Thus, the final susceptibility maps generated with a highly biased inventory appeared, with local exceptions (cf. “[Minor inventory biases locally affect modelling results](#)” section), remarkably similar to their references (e.g. compare Fig. 9a, j with 9i, r).

We argue that whenever there is a suspicion of an inventory bias (e.g. Brardinoni et al. 2003; Van Den Eckhaut et al. 2012; Petschko et al. 2016) and this incompleteness can systematically be described by a categorically scaled variable (e.g. land cover, administrative units), mixed-effects models (Bolker et al. 2009; Zuur et al. 2009) may be an appropriate choice. Subsequent predictions should be based on the models’ fixed-effects part alone. However, the separation of bias-related effects from landslide susceptibility-related effects may fail in the case of very high spatial agreements between a random intercept and other predictors (cf. example from the Molasse in “[Minor inventory biases locally affect modelling results](#)” section).

In the context of statistical landslide susceptibility modelling, mixed-effects models might additionally bear an up to now unexplored ability to tackle the problem of large and heterogeneous study areas (Petschko et al. 2014b). For instance, an inclusion of random regression coefficients would allow accounting for the effect that the relationship between a predictor (e.g. slope) and landslide occurrence differs between spatial units (e.g. lithologies). The application of generalized additive mixed models (GAMMs) may furthermore prove useful to additionally consider moderately non-linear associations (Zuur et al. 2009; Goetz et al. 2011).

The need for differentiated model evaluations

This study highlighted that an apparent high predictive performance of a model does not constitute proof that a realistic and geomorphically interpretable statistical landslide susceptibility model was generated. Under no circumstances, not even if compared to an unmodified inventory, did any model provide evidence of a poor prediction skill. AUROCs were constantly greater than 0.8. In this respect, we want to point out that it is known that this performance estimates are highly dependent on the study design and may considerably change when including easily classifiable areas (e.g. floodplains) (Lobo et al. 2008; Brenning 2012b). Thus, they do not have an absolute meaning and are not

comparable between different study areas. This is also why we think that frequently cited general guidelines for AUROC interpretations (e.g. in Hosmer and Lemeshow 2000) should not be used as a universal yardstick. Some presented models and maps of this study (e.g. Fig. 6f or Fig. 9f) are of little practical use and do clearly not provide an excellent representation of landslide susceptibility despite their AUROCs >0.85. In this context, we agree that an assessment of the prediction skill of a landslide susceptibility model is a crucial step, but not always sufficient (Lobo et al. 2008; Fressard et al. 2014; Steger et al. 2016a).

The findings further highlighted the need for differentiated evaluations (i.e. quantitative and expert-based) to gain insights into limitations and the reliability of landslide susceptibility analyses (Guzzetti et al. 2006b; Bell 2007; Demoulin and Chung 2007; Petschko et al. 2014b; Fressard et al. 2014; Steger et al. 2016a). In this sense, we agree that a preliminary in-depth inspection of input data should precede each landslide susceptibility analysis (Galli et al. 2008; Bell et al. 2012; Felicísimo et al. 2013; Santangelo et al. 2015). We consider field based cross-checks of inventory data and an in-depth exploratory data analysis as an essential supplement (Bell et al. 2012; Guillard and Zezere 2012; Petschko et al. 2016; Steger et al. 2016a). An estimation of modelled relationships and related uncertainties (i.e. confidences) might further provide evidence of model qualities and a potential bias propagation (Guzzetti et al. 2006b; Rossi et al. 2010; Petschko et al. 2014b; Steger et al. 2016a).

An evaluation of model transferabilities by means of repeated non-spatially assessed (CV) and spatially assessed (SCV) prediction skills may provide evidence of more or less consistent modelling results (e.g. Fig. 7d–f versus Fig. 10a–c). We observed that indications of spatially less transferable modelling results can be detected by an evaluation of the transferability index (Petschko et al. 2014b) and also by simply interpreting differences between AUROCs obtained by CV and SCV. Similar predictive performances (i.e. between CV and SCV; e.g. Fig. 7d–f) reveal that the respective models predicted all test sets equally well, independently of a spatial or non-spatial evaluation. In contrast, lower SCV results (e.g. Fig. 10a–c) provide evidence that the respective prediction models performed worse in a spatial context. This might be a standard case for real-world generated data sets (cf. Brenning 2005; Petschko et al. 2014b; Steger et al. 2016a), but especially for models generated with an inventory whose completeness differs substantially among larger geographical units (e.g. administrative units). We argue that an assessment of the spatial transferability of modelling results cannot only be useful to assess the ability of a model to predict landslides for a neighbouring area (Lombardo et al. 2014) but also to get indications of a potential spatially varying consistency of models within one area.

However, during this study, it became more and more evident that it is still the analyst who needs to assign meaning to the obtained numerical results. Thus, we finally argue that process-knowledge and an expert-based holistic interpretations of all results still remain an essential step towards meaningful statistical landslide susceptibility maps (Demoulin and Chung 2007; Cascini 2008; Zêzere et al. 2009; Fressard et al. 2014; Steger et al. 2016a, 2016b).

Practical recommendations

According to our findings, we recommend the following four-step procedure to tackle the problem of incomplete landslide

inventories when assessing landslide susceptibility by means of statistical classification techniques:

Firstly, we suggest accepting inventory errors as unavoidable. Without this realization, apparently well performing but highly distorted models of little explanatory power or practical use are more likely to follow.

Secondly, researchers should strive to gain insights into potential limitations of the present inventory data in order to assess possible implications for susceptibility modelling. This step should take into account details on the landslide data collection itself (e.g. type and resolution of mapping basis, spatial representation of landslides, spatial division of mapping responsibilities, mapping purpose) and also a broader geomorphic context (e.g. known causes of landslides and human impact in an area). These considerations should be supported by a profound literature review (e.g. which limitations are known when mapping from aerial photographs?), an exploratory data analysis (e.g. are there suspiciously high or low landslide densities within certain areas?) and field checks (e.g. are landslides missing in the inventory?).

Thirdly, modellers should consider potential limitations of the landslide inventory when adapting the modelling design. The aim should be to limit the propagation of inventory incompleteness into the final results. Here, we highly recommend avoiding predictors that directly relate to a suspected inventory error (e.g. land cover is likely to be able to directly describe a forest-related incompleteness of an inventory). Instead, the application of mixed-effects models may rather prove useful to reduce the impact of incompleteness. In this context, we suggest using strongly generalizing classifiers (e.g. GLMMs, GAMMs), because those models are likely to be less prone to overfit to errors originating from a landslide inventory (Steger et al. 2016b). The application of less flexible statistical models might additionally have the advantage of higher model transparency (e.g. via the estimation of ORs and confidence intervals) (Brenning 2012b; Goetz et al. 2015). This may further provide evidence of potential limitations and error propagation.

Fourthly, the results should continuously be evaluated by means of multiple quality criteria. Based on our results, we stress that an interpretation of one performance measure alone may lead to misleading conclusions. Modelled relationships may be evaluated by means of odds ratio estimation, while associated confidence intervals may provide insights into related uncertainties and an associated sensitivity to minor changes (and biases) in the data sets. We recommend assessing the predictive performance of a model by means of repeated partitioning techniques. We consider that an interpretation of CV with SCV results (i.e. median AUROC, interquartile ranges, transferability index) might be useful to expose inconsistent modelling results (Brenning 2012b). Finally, we want to further underline the importance for a steady geomorphological control over a purely data-driven treatment (Bell 2007; Demoulin and Chung 2007; Steger et al. 2016a), because domain experts need to interpret model results. Since landslide inventories might simultaneously be affected by positional inaccuracies and systematic incompleteness, we point to our complementary study that specifically focused on the propagation of inventory-based positional errors into statistical landslide susceptibility models (Steger et al. 2016b).

Conclusion

The study highlighted that the relation between biased landslide inventories and the results of statistical landslide susceptibility

models are complex and dependent on multiple aspects, such as the selection of predictors or the modelling approach. It was shown that high validation results, but distorted relationships and geomorphically implausible landslide susceptibility maps, can be obtained for models generated with systematically incomplete inventories. Most strikingly, we observed that an inclusion of a bias-describing predictor (e.g. land cover for a forest-related bias) favoured misleading predictive performance estimates and a direct bias propagation into subsequent models. However, an exclusion of such predictors led to confounded relationships between the landslide inventories and those predictors (e.g. slope) spatially related with these bias-describing variables. In this context, the application of mixed-effects logistic regression models reduced the influence of such confounders and enabled predictions that were less influenced by inventory bias.

We finally conclude that researchers should not only focus on predictive performances when modelling landslide susceptibility but also consider possible biases inherent in a landslide inventory. An in-depth evaluation of input data and modelling results, as well as an adaptation of model design, might prove valuable to reduce the impact of inventory errors on statistical landslide susceptibility models.

Acknowledgments

Open access funding provided by University of Vienna. We thank several Departments of the Provincial Government of Lower Austria for providing data: the Geological Survey, the Department of Spatial Planning and Regional Policy and the Department of Hydrology and Geoinformation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Ardizzone F, Cardinali M, Carrara A et al (2002) Impact of mapping errors on the reliability of landslide hazard maps. *Nat Hazards Earth Syst Sci* 2:3–14

Atkinson PM, Jiskoot H, Massari R, Murray T (1998) Generalized linear modelling in geomorphology. *Earth Surf Process Landf* 23:1185–1195

Atkinson PM, Massari R (2011) Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy. *Geomorphology* 130:55–64

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: linear mixed-effects models using Eigen and S4. R package version 1(7). <https://cran.r-project.org/web/packages/lme4/lme4.pdf>. Accessed 4 June 2016

Beguería S (2006) Validation and evaluation of predictive models in hazard assessment and risk management. *Nat Hazards* 37:315–329. doi:10.1007/s11069-005-5182-6

Bell R (2007) Lokale und regionale Gefahren- und Risikoanalyse gravitativer Massenbewegungen an der Schwäbischen Alb. Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn. <http://hss.ulb.uni-bonn.de/2007/1107/1107.htm>. Accessed 20 September 2016

Bell R, Petschko H, Röhrs M, Dix A (2012) Assessment of landslide age, landslide persistence and human impact using airborne laser scanning digital terrain models. *Geografiska Annaler: Series A, Physical Geography* 94:135–156. doi:10.1111/j.1468-0459.2012.00454.x

Bolker BM, Brooks ME, Clark CJ et al (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24:127–135

Brabb EE (1984) Innovative approaches to landslide hazard mapping. 4th International Symposium on Landslides, 16–21 September. Canadian Geotechnical Society, Toronto, pp 307–324

Brardinoni F, Slaymaker O, Hassan MA (2003) Landslide inventory in a rugged forested watershed: a comparison between air-photo and field survey data. *Geomorphology* 54:179–196. doi:10.1016/S0169-555X(02)00355-0

Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation. *Nat Hazards Earth Syst Sci* 5:853–862

Brenning A (2012a) Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package *sperrorest*. In: *Geoscience and Remote Sensing Symposium (IGARSS)*. 2012 I.E. International, Miami, pp 5372–5375

Brenning A (2012b) Improved spatial analysis and prediction of landslide susceptibility: practical recommendations. In: Eberhardt E, Froese C, Turner AK, Leroueil S (eds) *Landslides and engineered slopes, protecting society through improved understanding*. Taylor & Francis, Banff, Alberta, Canada 789–795

Brenning A, Schwinn M, Ruiz-Páez AP, Muenchow J (2015) Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Nat Hazards Earth Syst Sci* 15:45–57. doi:10.5194/nhess-15-45-2015

Budimir MEA, Atkinson PM, Lewis HG (2015) A systematic review of landslide probability mapping using logistic regression. *Landslides* 12:419–436. doi:10.1007/s10346-014-0550-5

Cascini L (2008) Applicability of landslide susceptibility and hazard zoning at different scales. *Eng Geol* 102:164–177. doi:10.1016/j.enggeo.2008.03.016

Chacón J, Irigaray C, Fernández T, Hamdouni RE (2006) Engineering geology maps: landslides and geographical information systems. *Bull Eng Geol Environ* 65:341–411. doi:10.1007/s10064-006-0064-z

Chung CJ, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30:451–472

Conoscenti C, Rotigliano E, Cama M et al (2016) Exploring the effect of absence selection on landslide susceptibility models: a case study in Sicily, Italy. *Geomorphology* 261:222–235

Corominas J, van Westen C, Frattini P et al (2013) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 73(2):209–263. doi:10.1007/s10064-013-0538-8

Cruden DM, Varnes DJ (1996) Landslide types and processes landslides: investigation and mitigation. TRB special report, 247. National Academy Press, Washington, pp 36–75

Demoulin A, Chung C-JF (2007) Mapping landslide susceptibility from small datasets: a case study in the Pays de Herve (E Belgium). *Geomorphology* 89:391–404. doi:10.1016/j.geomorph.2007.01.008

Dikau R, Brunsden D, Schrott L, Ibsen ML (1996) Landslide recognition: identification, movement and causes. Wiley, Chichester

Donati L, Turrini MC (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). *Eng Geol* 63:277–289

Eder A, Sotier B, Klebinder K, et al (2011) Hydrologische Bodenkenndaten der Böden Niederösterreichs (HydroBodNÖ), unpublished final report, Bundesamt für Wasserwirtschaft, Institut für Kulturtechnik und Bodenwasserhaushalt, Bundesforschungszentrum für Wald, Institut für Naturgefahren, Petzenkirchen, Innsbruck

Feliciísimo AM, Cuartero A, Remondo J, Quirós E (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides* 10:175–189. doi:10.1007/s10346-012-0320-1

Fell R, Corominas J, Bonnard C et al (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102:85–98. doi:10.1016/j.enggeo.2008.03.022

Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. *Eng Geol* 111:62–72. doi:10.1016/j.enggeo.2009.12.004

Fressard M, Thiery Y, Maquaire O (2014) Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d'Auge plateau hillslopes (Normandy, France). *Natural Hazards and Earth System Science* 14:569–588. doi:10.5194/nhess-14-569-2014

Galli M, Ardizzone F, Cardinali M et al (2008) Comparing landslide inventory maps. *Geomorphology* 94:268–289. doi:10.1016/j.geomorph.2006.09.023

Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput Geosci* 81:1–11. doi:10.1016/j.cageo.2015.04.007

Goetz JN, Guthrie RH, Brenning A (2011) Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology* 129:376–386. doi:10.1016/j.geomorph.2011.03.001

Greiving S, Mayer J, Pohl J, et al (2012) Kooperation zwischen Raumforschung und Raumplanungspraxis—diskutiert am Beispiel der Berücksichtigung von Hangrutschungsgefährdungen im Regionalplan Neckar-Alb. *RaumPlanung* 158:274–281

- Guillard C, Zezere J (2012) Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: the case of Loures Municipality. *Environ Manag* 50:721–735. doi:10.1007/s00267-012-9921-7
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216
- Guzzetti F, Galli M, Reichenbach P et al (2006a) Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Natural Hazards and Earth System Science* 6:115–131
- Guzzetti F, Mondini AC, Cardinali M et al (2012) Landslide inventory maps: new tools for an old problem. *Earth Sci Rev* 112:42–66. doi:10.1016/j.earscirev.2012.02.001
- Guzzetti F, Reichenbach P, Ardizzone F et al (2006b) Estimating the quality of landslide susceptibility models. *Geomorphology* 81:166–184. doi:10.1016/j.geomorph.2006.04.007
- Harp EL, Keefer DK, Sato HP, Yagi H (2011) Landslide inventories: the essential part of seismic landslide hazard analyses. *Eng Geol* 122:9–21
- Heckmann T, Gegg K, Gegg A, Becht M (2014) Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. *Natural Hazards and Earth System Science* 14:259–278. doi:10.5194/nhess-14-259-2014
- Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York
- Hovius N, Stark CP, Allen PA (1997) Sediment flux from a mountain belt derived by landslide mapping. *Geology* 25:231–234
- Huggett R (2007) Fundamentals of geomorphology. Routledge, London, New York
- Hussin HY, Zumpano V, Reichenbach P et al (2016) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology* 253:508–523. doi:10.1016/j.geomorph.2015.10.030
- Iovine GGR, Greco R, Gariano SL et al (2014) Shallow-landslide susceptibility in the Costa Viola mountain ridge (southern Calabria, Italy) with considerations on the role of causal factors. *Nat Hazards* 73:111–136. doi:10.1007/s11069-014-1129-0
- Jacobs L, Dewitte O, Poesen J, et al (2016) Landslide characteristics and spatial distribution in the Rwenzori Mountains, Uganda. *Journal of African Earth Sciences*. doi: 10.1016/j.jafrearsci.2016.05.013
- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17:145–151. doi:10.1111/j.1466-8238.2007.00358.x
- Lombardo L, Cama M, Maerker M, Rotigliano E (2014) A test of transferability for landslides susceptibility models under extreme climatic events: application to the Messina 2009 disaster. *Nat Hazards* 74:1951–1989. doi:10.1007/s11069-014-1285-2
- Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. *Earth Surf Process Landf* 29:687–711. doi:10.1002/esp.1064
- Muenchow J, Brenning A, Richter M (2012) Geomorphic process rates of landslides along a humidity gradient in the tropical Andes. *Geomorphology* 139–140:271–284. doi:10.1016/j.geomorph.2011.10.029
- Petschko H, Bell R, Glade T (2016) Effectiveness of visually analyzing LiDAR DTM derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling. *Landslides* 13(5):857–872. doi:10.1007/s10346-015-0622-1
- Petschko H, Bell R, Glade T (2014a) Relative age estimation at landslide mapping on LiDAR derivatives: revealing the applicability of land cover data in statistical susceptibility modelling. In: Sassa K, Canuti P, Yin Y (eds) *Landslide science for a safer geoenvironment*. Springer International Publishing, New York City, pp 337–343
- Petschko H, Brenning A, Bell R et al (2014b) Assessing the quality of landslide susceptibility maps—case study Lower Austria. *Natural Hazards and Earth System Science* 14:95–118. doi:10.5194/nhess-14-95-2014
- Poli S, Sterlacchini S (2007) Landslide representation strategies in susceptibility studies using weights-of-evidence modeling technique. *Nat Resour Res* 16:121–134. doi:10.1007/s11053-007-9043-8
- Core Team R (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Regmi NR, Giardino JR, McDonald EV, Vitek JD (2014) A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. *Landslides* 11(2):247–262. doi:10.1007/s10346-012-0380-2
- Reichenbach P, Busca C, Mondini AC, Rossi M (2014) The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). *Environ Manag* 54:1372–1384. doi:10.1007/s00267-014-0357-0
- Remondo J, González A, De Terán JRD et al (2003) Validation of landslide susceptibility maps: examples and applications from a case study in Northern Spain. *Nat Hazards* 30:437–449
- Rickli C, Zürcher K, Frey W, Lüscher P (2002) Wirkungen des Waldes auf oberflächennahe Rutschprozesse. *Schweiz Z Forstwes* 153(11):437–445
- Rossi M, Guzzetti F, Reichenbach P et al (2010) Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114:129–142. doi:10.1016/j.geomorph.2009.06.020
- Santangelo M, Marchesini I, Bucci F et al (2015) An approach to reduce mapping errors in the production of landslide inventory maps. *Natural Hazards and Earth System Science* 15:2111–2126. doi:10.5194/nhess-15-2111-2015
- Schwenk H (1992) Massenbewegungen in Niederösterreich 1953–1990. *Jahrb. Geol. Bundesanst. Geologische Bundesanstalt, Wien*, pp 597–660
- Skoda G, Lorenz P (2007) Mean annual precipitation—modelled with uncorrected data. Digital version of the hydrological atlas of Austria (DigHAO 2007). <https://iwfw.boku.ac.at/haof/>. Accessed 22 July 2016
- Soeters R, Van Westen CJ (1996) Slope instability recognition, analysis and zonation. In: Turner AK, Schuster RL (eds) *Landslides: investigation and mitigation*. Transportation Research Board National Research Council, Washington, D.C, pp 129–177
- Steger S, Bell R, Petschko H, Glade T (2015) Evaluating the effect of modelling methods and landslide inventories used for statistical susceptibility modelling. In: Lollino G, Giordan D, Crosta GB et al (eds) *Engineering geology for society and territory—volume 2*. Springer International Publishing, New York City, pp 201–204
- Steger S, Brenning A, Bell R et al (2016a) Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262:8–23. doi:10.1016/j.geomorph.2016.03.015
- Steger S, Brenning A, Bell R, Glade T (2016b) The propagation of inventory-based positional errors into statistical landslide susceptibility models. *Nat Hazards Earth Syst Sci Discuss* 16:2729–2745. doi:10.5194/nhess-16-2729-2016
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293
- Szkló M, Nieto FJ (2014) Epidemiology: beyond the basics, 3rd edn. Jones & Bartlett Learning, Burlington, Mass
- Van Den Eeckhaut M, Hervás J, Jaedicke C et al (2012) Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* 9:357–369. doi:10.1007/s10346-011-0299-z
- Van Den Eeckhaut M, Marre A, Poesen J (2010) Comparison of two landslide susceptibility assessments in the Champagne-Ardenne region (France). *Geomorphology* 115:141–155. doi:10.1016/j.geomorph.2009.09.042
- Van Den Eeckhaut M, Vanwalleghe T, Poesen J et al (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). *Geomorphology* 76:392–410. doi:10.1016/j.geomorph.2005.12.003
- Van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Eng Geol* 102:112–131. doi:10.1016/j.enggeo.2008.03.010
- Vorpahl P, Elsenbeer H, Märker M, Schröder B (2012) How can statistical models help to determine driving factors of landslides? *Ecol Model* 239:27–39. doi:10.1016/j.ecolmodel.2011.12.007
- Wessely G, Draxler I, Gangl G et al (2006) *Geologie der österreichischen Bundesländer—Niederösterreich*. Geologische Bundesanstalt, Wien
- Xu C, Xu X, Shyu JBH et al (2015) Landslides triggered by the 20 April 2013 Lushan, China, Mw 6.6 earthquake from field investigations and preliminary analyses. *Landslides* 12:365–385. doi:10.1007/s10346-014-0546-1
- Zezere J, Henriques CS, Garcia RAC, Piedade A (2009) Effects of landslide inventories uncertainty on landslide susceptibility modelling. *RISKam Geographical Research Centre*. University of Lisbon, Portugal, Lisbon http://eost.u-strasbg.fr/omiv/Landslide_Processes_Conference/Zezere_et_al.pdf. Accessed 29 July 2016
- Zuur A, Ieno EN, Walker N et al (2009) Mixed effects models and extensions in ecology with R, 2009th edn. Springer, New York, NY

S. Steger (✉) · R. Bell · T. Glade

Department of Geography and Regional Research,
University of Vienna,
Universitätsstraße 7, 1010, Vienna, Austria
e-mail: stefan.steger@univie.ac.at

A. Brenning

Department of Geography,
Friedrich-Schiller University Jena,
Loebdergraben 32, 07743, Jena, Germany